

# Unveiling Bias Compensation in Turbo-Based Algorithms for (Discrete) Compressed Sensing

Susanne Sparrer, Robert F.H. Fischer

Institute of Communications Engineering, Ulm University, 89081 Ulm, Germany

Email: susanne.sparrer@uni-ulm.de, robert.fischer@uni-ulm.de

**Abstract**—In Compressed Sensing, a real-valued sparse vector has to be recovered from an underdetermined system of linear equations. In many applications, however, the elements of the sparse vector are drawn from a finite set. Adapted algorithms incorporating this additional knowledge are required for the discrete-valued setup. In this paper, turbo-based algorithms for both cases are elucidated and analyzed from a communications engineering perspective, leading to a deeper understanding of the algorithm. In particular, we gain the intriguing insight that the calculation of extrinsic values is equal to the unbiasing of a biased estimate and present an improved algorithm.

## I. INTRODUCTION

In many communication scenarios, the transmitted vector  $\mathbf{x} \in \mathbb{R}^{L \times 1}$  is *sparse*, i.e., only  $s$  elements are non-zero.<sup>1</sup> The receive vector  $\mathbf{y} \in \mathbb{R}^{K \times 1}$ ,  $s \ll K < L$ , is corrupted by i.i.d. zero-mean Gaussian noise  $\mathbf{n}$ , with variance  $\sigma_n^2$  per component. The channel can then be modeled by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad (1)$$

where  $\mathbf{A} \in \mathbb{R}^{K \times L}$  corresponds to the channel matrix in terms of communications engineering. Since  $K < L$ , a sparse vector has to be recovered from an underdetermined system of linear equations, a problem which is known as *Compressed Sensing* (CS) [1]. In many communication scenarios, however, the non-zero elements are not real-valued but drawn from a finite set  $\mathcal{C}$ . If the sparsity is fixed, the problem to be solved is given by ( $\mathcal{C}_0 \stackrel{\text{def}}{=} \mathcal{C} \cup \{0\}$ )

$$\hat{\mathbf{x}} = \underset{\tilde{\mathbf{x}} \in \mathcal{C}_0^L}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}\|_2^2 \quad \text{s.t.} \quad \|\tilde{\mathbf{x}}\|_0 = s. \quad (2)$$

There are many fields of digital communication in which this problem of estimating a *discrete-valued sparse* vector from an *underdetermined* system of linear equations is present, such as, e.g., sensor networks, where a fusion center with  $K$  antennas has to reconstruct which of the  $L$  low-activity sensors have currently been active, and which data has been transmitted by them [2]. Further applications are peak-to-average power reduction in orthogonal frequency-division multiplexing [3],

This work was supported by Deutsche Forschungsgemeinschaft (DFG) under grant FI 982/8-1.

<sup>1</sup>Notation:  $\|\cdot\|_p$  denotes the  $\ell_p$  norm.  $\mathbf{A}_{(l,m)} = A_{l,m}$  is the element in the  $l^{\text{th}}$  row and  $m^{\text{th}}$  column of  $\mathbf{A}$ .  $\mathbf{A}^T$  and  $\mathbf{A}^{-1}$  denote the transpose and the inverse of  $\mathbf{A}$ , respectively.  $\mathbf{diag}(\mathbf{a})$  denotes a diagonal matrix of appropriate size with entries of the vector  $\mathbf{a}$  as diagonal elements.  $\mathbf{diag}(\mathbf{A})$  denotes a diagonal matrix with the same diagonal elements as  $\mathbf{A}$ .  $\mathbf{I}$  is the identity matrix.  $\mathcal{Q}_C(\cdot)$ : element-wise quantization w.r.t. a given alphabet  $\mathcal{C}$ .  $\mathbb{E}\{\cdot\}$ : element-wise expectation.  $\text{Var}\{\cdot\}$ : Variance.  $\mathcal{N}(m, v)$ : Gaussian distribution with mean  $m$  and variance  $v$ .  $f_X(x)$ : probability density function of random variable  $x$ .

the detection of pulse-width-modulated signals in radar applications [4], code-book excited linear prediction (CELP) source coding [5], and Compressed-Sensing-based cryptography [6].

There is a tremendously wide range of algorithms solving the standard continuous-valued CS problem, such as, amongst others, Orthogonal Matching Pursuit (OMP) [7], Iterative Hard Thresholding (IHT) [8], and Iterative Soft Thresholding (IST) [9]. Although the standard CS problem is non-convex due to its sparsity constraint, it can be relaxed to a  $\ell_1$ -based problem, which can efficiently be solved by the simplex algorithm or interior point methods [10].

In the case of *discrete Compressed Sensing*, however, additional information, i.e., the knowledge that the elements of the sparse vector are from a finite set, is available and has to be taken into account adequately. The estimation of a discrete-valued vector has combinatorial complexity in general, and hence discrete CS is non-convex, even if the constraint in problem (2) was relaxed to an  $\ell_1$ -based one, which could be solved by extensions of the simplex algorithm [10]. Unfortunately, these algorithms have a prohibitively high computational complexity.

Some algorithms for the solution of problem (2) have been proposed over the last few years. Besides the most obvious approach of a standard CS algorithm with subsequent quantizer [11], the quantization can be included inside OMP [13], which equals the so-called model-based Compressed Sensing [12] if it is applied to discrete CS. This algorithm has been further improved by the application of a method which preserves reliability information [13]. Another improved variant of OMP has been introduced in [14], where a minimum mean-squared error estimator has been applied.

Other algorithms for the CS problem are related to well-known channel decoding algorithms, e.g., the approximate message-passing (AMP) algorithm [15], [16], which is derived from the message-passing algorithm [17] and which can be easily adapted to cases (such as discrete CS) where information on the a-priori distribution of the sparse vector is available (Bayesian AMP, BAMP) [15]. Moreover, the knowledge from channel coding has also been used for the optimization of measurement matrices with adapted specialized recovery algorithms, cf., e.g., [18], [19]. The drawback of these approaches is that the restrictions on the measurement matrix limit the range of applications.

In [20], [21], an approach which is based on the turbo-principle has been proposed. It has been simplified, generalized, and adapted to the discrete setup in [22]. In this

paper, this algorithm is further improved, thereby including knowledge from the field of digital communications into CS. The aim of this paper is to gain a profound understanding of the algorithm. An analysis of the approaches leads to intriguing insights into the algorithm, and especially into the important topic of bias compensation. Note that the results hold for standard CS and for discrete CS.

The paper is organized as follows. In Sec. II, the improved algorithm is introduced. The analysis is given in Sec. III. In Sec. IV, the performance of the algorithms is compared, followed by brief conclusions in Sec. V.

## II. TURBO SIGNAL RECOVERY

In [20], [21], an iterative algorithm for the estimation of complex-valued sparse vectors has been presented. This algorithm, which has been denoted as Turbo Signal Recovery (TSR) (or Turbo Compressed Sensing), has been restricted to one type of measurement matrices in the original work. In [22], the algorithm has been generalized to a wider range of matrix construction, which allowed the comparison of TSR with other algorithms. Furthermore, the notation of the algorithm has been simplified. In this section, after a short explanation of this modified algorithm, an improved version is introduced.

### A. Approximate LMMSE TSR

In the generalized TSR algorithm, the measurement matrix is assumed to be given by  $\mathbf{A} = \mathbf{U}\mathbf{C}$ , where  $\mathbf{U}$  is a random part of a unitary matrix, and  $\mathbf{C} = \text{diag}([c_1, \dots, c_L])$  is a scaling matrix. In the standard CS setup where the *column* vectors of  $\mathbf{A}$  are assumed to be normalized to unit length, the scaling factors calculate to  $c_i^2 = 1 / \sum_{j=1}^K U_{j,i}^2$ .

The sparse vector is estimated in an iterative fashion, where each iteration consists of two main parts: In the first part, all elements of the vector are estimated jointly by a linear estimator aiming to keep the Euclidean estimation error small, thereby ignoring the sparsity and the alphabet constraint. In the second part, the constraints which have been disregarded in the first part are taken into account, which leads to a non-linear estimator which generates so-called soft values [23]. The pseudocode of this algorithm is given in Alg. 1, Variant A, i.e., only the lines tagged by an “A” are active.

The authors of the original TSR algorithm denote the first step as linear minimum mean-square error (LMMSE) estimation [20]. In this step, the sparse vector is estimated by (cf. Alg. 1, Line 3)

$$\mathbf{x}_M^{\text{post}} = \mathbf{x}_M^{\text{pri}} + \frac{\bar{c}^2 \sigma_{M,\text{pri}}^2}{\bar{c}^2 \sigma_{M,\text{pri}}^2 + \sigma_n^2} \cdot \mathbf{C}^{-1} \mathbf{U}^T (\mathbf{y} - \mathbf{A} \mathbf{x}_M^{\text{pri}}), \quad (3)$$

where  $\mathbf{x}_M^{\text{pri}}$  is a prior estimate (from the previous step),  $\sigma_{M,\text{pri}}^2$  is the variance of the estimation error of this prior estimate, and  $\bar{c}^2 = \frac{1}{L} \sum_{i=1}^L c_i^2$  is the average scaling factor. All variables of this MMSE-estimation-step are marked by the index “M”.

In [22], it has been shown that with  $\mathbf{C} \approx \bar{c} \mathbf{I}$ , (3) can be approximated by

$$\mathbf{x}_M^{\text{post}} \approx \mathbf{x}_M^{\text{pri}} + \frac{\sigma_{M,\text{pri}}^2}{\bar{c}^2 \sigma_{M,\text{pri}}^2 + \sigma_n^2} \cdot \mathbf{A}^T (\mathbf{y} - \mathbf{A} \mathbf{x}_M^{\text{pri}}). \quad (4)$$

Note that (when ignoring the scaling factor) this term corresponds to the first step in the well-known IHT algorithm [8], which can, on the one hand, be interpreted as one step of the gradient descent method, but, on the other hand, as correlation-based estimation, since  $\mathbf{x}$  is estimated based on the correlation between the residual and the column vectors of  $\mathbf{A}$ . From a communications engineering point of view, it corresponds to the application of a matched filter. Hence, although claimed otherwise by the authors in [20], the first step is not an LMMSE estimation.

In the second step, soft values are calculated, which are the expected value of  $x$  conditioned to the prior estimate from the first step and the a-priori distribution of  $x$  (cf. Alg. 1, Line 8). Note that, in contrast to the joint estimation in the first step, this calculation is performed for each element of the sparse vector individually. Since this approach takes the a-priori distribution of  $x$  into account, it depends on the alphabet; an adaptation to any alphabet is straightforward. This approach is also used in other algorithms for (discrete) CS, cf., e.g., [15], [22], [24], [13]. All variables of the second (soft-value calculating) step are indicated by the index “S”.

This algorithm is denominated as TSR/Q in the following, where the trailing “Q” emphasizes the final quantization step which has to be performed in order to restrict the estimate to the discrete alphabet.

---

**Alg. 1**  $\hat{\mathbf{x}} = \text{recover}(\mathbf{y}, \mathbf{U}, \mathbf{C}, \sigma_n^2, s, c_0)$

---

Variants: A: TSR/Q, B: TMS/Q

---

1AB:  $\mathbf{x}_M^{\text{pri}} = \mathbf{0}$ ,  $\sigma_{M,\text{pri}}^2 = s/L$ ,  $\mathbf{A} = \mathbf{U}\mathbf{C}$

2AB: **while** stopping criterion not met {

// MMSE estimation

3A :  $\mathbf{x}_M^{\text{post}} = \mathbf{x}_M^{\text{pri}} + \frac{\bar{c}^2 \sigma_{M,\text{pri}}^2}{\bar{c}^2 \sigma_{M,\text{pri}}^2 + \sigma_n^2} \cdot \mathbf{C}^{-1} \mathbf{U}^T (\mathbf{y} - \mathbf{A} \mathbf{x}_M^{\text{pri}})$

3 B:  $\mathbf{x}_M^{\text{post}} = \mathbf{x}_M^{\text{pri}} + \sigma_{M,\text{pri}}^2 \mathbf{A}^T (\sigma_{M,\text{pri}}^2 \mathbf{A} \mathbf{A}^T + \sigma_n^2 \mathbf{I})^{-1} (\mathbf{y} - \mathbf{A} \mathbf{x}_M^{\text{pri}})$

4A :  $\sigma_{M,\text{post}}^2 = \sigma_{M,\text{pri}}^2 \cdot \left( 1 - \frac{K}{L} \frac{\bar{c}^2 \sigma_{M,\text{pri}}^2}{\bar{c}^2 \sigma_{M,\text{pri}}^2 + \sigma_n^2} \right)$

4 B:  $\mathbf{K} = \sigma_{M,\text{pri}}^2 \mathbf{A}^T (\sigma_{M,\text{pri}}^2 \mathbf{A} \mathbf{A}^T + \sigma_n^2 \mathbf{I})^{-1} \mathbf{A}$

5 B:  $\sigma_{M,\text{post}}^2 = \sigma_{M,\text{pri}}^2 \cdot \left( 1 - \frac{1}{L} \sum_{i=1}^L K_{i,i} \right)$

6AB:  $\sigma_{S,\text{pri}}^2 = \sigma_{M,\text{ext}}^2 = \left( \frac{1}{\sigma_{M,\text{post}}^2} - \frac{1}{\sigma_{M,\text{pri}}^2} \right)^{-1}$

7AB:  $\mathbf{x}_S^{\text{pri}} = \mathbf{x}_M^{\text{ext}} = \sigma_{M,\text{ext}}^2 \left( \frac{\mathbf{x}_M^{\text{post}}}{\sigma_{M,\text{post}}^2} - \frac{\mathbf{x}_M^{\text{pri}}}{\sigma_{M,\text{pri}}^2} \right)$

// Soft feedback

8AB:  $x_{S,i}^{\text{post}} = \mathbb{E}\{x_i | x_{S,i}^{\text{pri}}\} = \mathcal{W}(x_{S,i}^{\text{pri}}, \sigma_{S,\text{pri}}^2, s)$

9AB:  $\sigma_{S,\text{post}}^2 = \frac{1}{L} \sum_{i=1}^L \text{Var}\{x_i | x_{S,i}^{\text{pri}}\}$

10AB:  $\sigma_{M,\text{pri}}^2 = \sigma_{S,\text{ext}}^2 = \left( \frac{1}{\sigma_{S,\text{post}}^2} - \frac{1}{\sigma_{S,\text{pri}}^2} \right)^{-1}$

11AB:  $\mathbf{x}_M^{\text{pri}} = \mathbf{x}_S^{\text{ext}} = \sigma_{S,\text{ext}}^2 \left( \frac{\mathbf{x}_S^{\text{post}}}{\sigma_{S,\text{post}}^2} - \frac{\mathbf{x}_S^{\text{pri}}}{\sigma_{S,\text{pri}}^2} \right)$

12AB: }

13AB:  $\hat{\mathbf{x}} = \mathcal{Q}_{c_0}(\mathbf{x}_S^{\text{post}})$

---

### B. Exact LMMSE TSR

After the introduction of TSR/Q in the previous section, an improved version of the algorithm is introduced in this section. To this end, the true linear MMSE estimate for  $\mathbf{x}$  is derived.

We assume the channel model (1), i.e.,  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ , and want an estimate  $\mathbf{x}_M^{\text{post}}$  of  $\mathbf{x}$  that minimizes the expectation of the squared error given prior knowledge on  $\mathbf{x}$ . In general, the linear MMSE estimator is given by [25], [22]

$$\mathbf{x}_M^{\text{post}} = \mathbf{x}_M^{\text{pri}} + \Phi_{dd} \mathbf{A}^T \left( \mathbf{A} \Phi_{dd} \mathbf{A}^T + \sigma_n^2 \mathbf{I} \right)^{-1} (\mathbf{y} - \mathbf{A} \mathbf{x}_M^{\text{pri}}),$$

where  $\mathbf{d}$  is the error vector if  $\mathbf{x}_M^{\text{pri}} = \mathbf{x} + \mathbf{d}$  is written as a noisy variant of  $\mathbf{x}$ , and  $\Phi_{dd} = \mathbb{E}\{\mathbf{d}\mathbf{d}^T\}$  is its correlation matrix. If we assume that the elements of  $\mathbf{d}$  are uncorrelated with variance  $\sigma_{M,\text{pri}}^2$ , the estimation can be simplified to

$$\mathbf{x}_M^{\text{post}} = \mathbf{x}_M^{\text{pri}} + \mathbf{A}^T \left( \mathbf{A} \mathbf{A}^T + \frac{\sigma_n^2}{\sigma_{M,\text{pri}}^2} \mathbf{I} \right)^{-1} (\mathbf{y} - \mathbf{A} \mathbf{x}_M^{\text{pri}}).$$

The calculation of the soft values in the second step is equal to the one in TSR.

This turbo-based algorithm, combining MMSE estimation and soft feedback, is denoted by TMS/Q. The pseudocode of this algorithm is given in Alg. 1, Variant B.

A comparison with the linear estimation in TSR/Q (cf. (4)) shows that TSR/Q can be seen as a simplified version of TMS/Q, valid for  $\mathbf{A} \mathbf{A}^T \approx \bar{c}^2 \mathbf{I}$ . Note that TMS/Q holds for all types of measurement matrices, the restriction that  $\mathbf{A}$  is a (scaled) part of a unitary matrix, which was assumed in the original paper on TSR [20] and also for TSR/Q, is not required anymore.

### III. DISCUSSION OF THE EXTRINSICS CALCULATION

The TMS algorithm is based on the turbo principle, which is well-known in channel coding [26]. The general idea is that two decoders (A and B in general, LMMSE estimator and soft value calculation here) exist, which alternately decode or estimate the signal and which exchange information about the current results. Each decoder takes into account the information from the other decoder (the so-called priors, i.e.,  $\mathbf{x}_M^{\text{pri}}$  &  $\mathbf{x}_S^{\text{pri}}$ ) and calculates a new estimate, the so-called posteriors ( $\mathbf{x}_M^{\text{post}}$  and  $\mathbf{x}_S^{\text{post}}$ ). In order not to hand back the information to the other decoder which has been received from it, this prior information has to be removed from the estimate (resulting in the so-called extrinsics, i.e.,  $\mathbf{x}_M^{\text{ext}}$  &  $\mathbf{x}_S^{\text{ext}}$ ) before passing it to the other decoder. The *extrinsic values* are calculated by (cf. Line 7 & 11, Alg. 1)

$$\mathbf{x}_S^{\text{pri}} = \mathbf{x}_M^{\text{ext}} = \sigma_{M,\text{ext}}^2 \left( \frac{\mathbf{x}_M^{\text{post}}}{\sigma_{M,\text{post}}^2} - \frac{\mathbf{x}_M^{\text{pri}}}{\sigma_{M,\text{pri}}^2} \right), \quad (5)$$

with corresponding average variances

$$\sigma_{S,\text{pri}}^2 = \sigma_{M,\text{ext}}^2 = \left( \frac{1}{\sigma_{M,\text{post}}^2} - \frac{1}{\sigma_{M,\text{pri}}^2} \right)^{-1}. \quad (6)$$

This principle is shown in the upper part of Fig. 1.

In the following, both blocks of the algorithm are analyzed from a different point of view, leading to interesting insights.

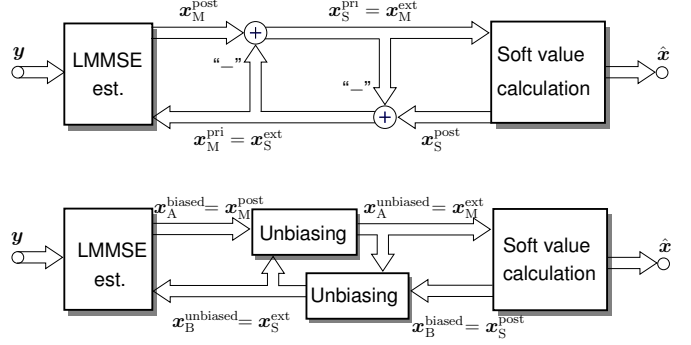


Fig. 1: Block diagram of TMS, interpreted from a turbo perspective (upper part), and interpreted from a signal theory perspective (lower part).

#### A. First Part

If discrete-valued signals are to be recovered, it is important that the diagonal elements of the end-to-end cascade for the estimation of  $\mathbf{x}$  are equal to 1, otherwise a *bias* is present. In the MMSE case, the cascade is given by

$$\mathbf{K} = [K_{i,j}] = \mathbf{A}^T \left( \mathbf{A} \mathbf{A}^T + \frac{\sigma_n^2}{\sigma_{M,\text{pri}}^2} \mathbf{I} \right)^{-1} \mathbf{A}, \quad (7)$$

with diagonal elements smaller than one. In order to compensate for the bias, the estimates have to be scaled by the inverse of the diagonal elements of  $\mathbf{K}$ . Given the *biased* estimate  $\mathbf{x}_M^{\text{pri}}$ , the  $i^{\text{th}}$  *unbiased* element  $x_{U,i}$  is estimated by

$$\begin{aligned} x_{U,i} &= x_{M,i}^{\text{pri}} + \frac{1}{K_{i,i}} \cdot (x_{M,i}^{\text{post}} - x_{M,i}^{\text{pri}}) \\ &= \frac{1}{K_{i,i}} \cdot x_{M,i}^{\text{post}} - \left( \frac{1}{K_{i,i}} - 1 \right) \cdot x_{M,i}^{\text{pri}}. \end{aligned} \quad (8)$$

With the prior error variance  $\sigma_{M,\text{pri}}^2$ , the *biased* a-posteriori error variance of the  $i^{\text{th}}$  element is given by [25]

$$\begin{aligned} \sigma_{M,\text{post},i}^2 &= \sigma_{M,\text{pri}}^2 - \sigma_{M,\text{pri}}^2 \cdot \left[ \mathbf{A}^T \left( \mathbf{A} \mathbf{A}^T + \frac{\sigma_n^2}{\sigma_{M,\text{pri}}^2} \mathbf{I} \right)^{-1} \mathbf{A} \right]_{(ii)} \\ &= \sigma_{M,\text{pri}}^2 \cdot (1 - K_{i,i}), \end{aligned} \quad (9)$$

and the *unbiased* error variance  $\sigma_{U,i}^2$  calculates to [22]

$$\sigma_{U,i}^2 = \sigma_{M,\text{pri}}^2 \cdot \left( \frac{1}{K_{i,i}} - 1 \right). \quad (10)$$

Combining (8)–(10), it follows

$$x_{U,i} = \sigma_{U,i}^2 \cdot \left( \frac{x_{M,i}^{\text{post}}}{\sigma_{M,\text{post},i}^2} - \frac{x_{M,i}^{\text{pri}}}{\sigma_{M,\text{pri}}^2} \right) \quad (11)$$

$$= x_{M,i}^{\text{ext}} \quad (12)$$

$$\sigma_{U,i}^2 = \sigma_{M,\text{ext},i}^2, \quad (13)$$

which is equal to the calculation of the extrinsic values (cf. Line 7, Alg. 1). Hence, the extrinsic calculation corresponds to unbiasing of the estimate, and thus, when using unbiased MMSE estimates, (inherently) extrinsic information is considered.

The unbiasing operation can also be interpreted from a third point of view. Given an observation  $o$ , the probability density function (pdf) of the biased a-posteriori estimate is given by

$$f_X(x|o) \sim \mathcal{N}(x_{M,i}^{\text{post}}, \sigma_{M,\text{post}}^2), \quad (14)$$

which corresponds to the *backward* channel model, cf. Fig. 2, upper part. Note that in this model, as in any MMSE solution, the error is uncorrelated to the observation  $o$ . After the bias compensation, however, the error is uncorrelated to the unbiased estimate [27], which leads to the *forward* channel model which is shown in the lower part of Fig. 2. In this case, the density of interest is  $f_O(o|x)$ . With Bayes' theorem it holds

$$f_O(o|x) = f_O(o) \frac{f_X(x|o)}{f_X(x)}, \quad (15)$$

where  $f_O(o)$  can be considered as a constant for a given  $o$ . If all variables are Gaussian distributed, it follows

$$f_O(o|x) \sim \frac{\mathcal{N}(x_{M,i}^{\text{post}}, \sigma_{M,\text{post}}^2)}{\mathcal{N}(x_{M,i}^{\text{pri}}, \sigma_{M,\text{pri}}^2)}, \quad (16)$$

with the expected value and variance [28]

$$x_{M,i}^{\text{ext}} = \mathbb{E}\{o|x_{M,i}^{\text{pri}}\} = \sigma_{M,\text{ext}}^2 \cdot \left( \frac{x_{M,i}^{\text{post}}}{\sigma_{M,\text{post}}^2} - \frac{x_{M,i}^{\text{pri}}}{\sigma_{M,\text{pri}}^2} \right) \quad (17)$$

$$\sigma_{M,\text{ext}}^2 = \text{Var}\{o|x_{M,i}^{\text{pri}}\} = \left( \frac{1}{\sigma_{M,\text{post}}^2} - \frac{1}{\sigma_{M,\text{pri}}^2} \right)^{-1}, \quad (18)$$

which are again the equations used for the extrinsic calculation in TMS. Thus, the *biased* MMSE estimate is equal to the *a-posteriori* value, which corresponds to the *backward* channel model. The *unbiased* MMSE estimate is equal to the *extrinsic* value, and thus the *forward* channel model.

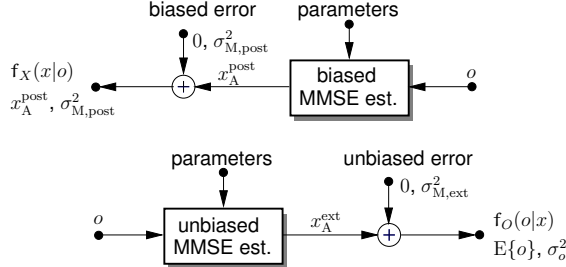


Fig. 2: Channel models in the estimation process.

### B. Second Part

In the second part of TMS, soft values are calculated. Since the estimator is non-linear, the calculation of the bias is not as obvious as for the first step. However, every non-linear memoryless device can be written as a linear scaling plus an additive estimation error which is uncorrelated to the linear part [29]. In general, any prior estimate  $z$  of the transmitted value  $x$  (with  $z = x_{S,i}^{\text{pri}}$  in our case) can be written as noisy variant of  $x$ , i.e.,  $z = x + e$ , where we assume the error to be Gaussian  $e \sim \mathcal{N}(0, \sigma_e^2)$ , with  $\sigma_e^2 = \sigma_{S,\text{pri}}^2$ , and independent from  $x$ . The soft value of  $z$  can be written as a function  $x_{S,i}^{\text{post}} = g(z) = \mathbb{E}\{x_i|z\}$ ,  $z = x_{S,i}^{\text{pri}}$ , which in turn is linearized as [29]

$$z - g(z) = k_E e + n_E, \quad (19)$$

TABLE I: Comparison of the steps.

	First step	Second step
Processing	vectorwise (joint)	symbolwise (individual)
Assumption	$x$ Gaussian distributed	sparsity $s$ , $x \in \mathcal{C}_0$
Estimation	linear/affine	non-linear
Estimated variable	signal $x$	error $e$

where the scaling factor

$$k_E = \frac{\mathbb{E}_{ZE}\{(z - g(z))e\}}{\sigma_e^2} \quad (20)$$

is chosen such that the error  $n_E$  has minimum variance. Note that, in contrast to the first step where we estimate the *signal*  $x$ , the *error*  $e$  is estimated, with which in turn  $x$  can be calculated. The estimated error after the soft feedback is given by  $x + e - g(x + e)$ , and hence

$$k_E = \frac{\mathbb{E}_{XE}\{(x + e - g(x + e))e\}}{\sigma_e^2} \quad (21)$$

$$= \frac{\sigma_e^2 - \mathbb{E}\{g(x + e)e\}}{\sigma_e^2}. \quad (22)$$

It can be shown via integration by parts that for any distribution of  $x$ ,  $\mathbb{E}\{g(x + e)e\} = \mathbb{E}\{g(x_{S,i}^{\text{pri}})e\} = \sigma_{S,\text{post}}^2 = \mathbb{E}\{x_i^2|x_{S,i}^{\text{pri}}\} - (\mathbb{E}\{x_i|x_{S,i}^{\text{pri}}\})^2$  holds if the *error* is Gaussian distributed.<sup>2</sup> Hence, taking  $\sigma_e^2 = \sigma_{S,\text{pri}}^2$  into account,  $k_E = \frac{\sigma_{S,\text{pri}}^2 - \sigma_{S,\text{post}}^2}{\sigma_{S,\text{pri}}^2}$ , and with the resulting unbiasing factor  $1/k_E$ , the unbiased estimate calculates to  $x_{S,i}^{\text{ext}} = z - \frac{1}{k_E} \cdot (z - g(z)) = x_{S,i}^{\text{pri}} - \frac{\sigma_{S,\text{pri}}^2}{\sigma_{S,\text{pri}}^2 - \sigma_{S,\text{post}}^2} \cdot (x_{S,i}^{\text{pri}} - x_{S,i}^{\text{post}})$  and  $\sigma_{S,\text{ext}}^2 = \mathbb{E}\{(x - x_{S,i}^{\text{ext}})^2\}$ , which, after straightforward modifications, results in the same unbiasing formulas as in the first step. Hence, as in the first step, the extrinsic calculation in TMS is equal to unbiasing, i.e., the change from the backward channel model to the forward one. The block diagram of TMS when it is interpreted from this unbiasing perspective is given in the lower part of Fig. 1

A comparison of both steps is given in Table I. While the first step performs a vectorwise linear estimation of  $x$  (implicitly assuming that  $x$  is Gaussian distributed), a non-linear symbolwise estimator for the error  $e$  is applied in the second step, thereby taking into account the sparsity and alphabet constraint.

### IV. SIMULATION RESULTS

In this section, numerical results of the new algorithm are shown and compared to the ones of established algorithms. Two different channel matrices are used. First,  $\mathbf{A}$  is constructed as random part of a random orthogonal matrix. Second,  $\mathbf{A}$  is a random Gaussian matrix. In both cases, the columns are normalized to unit length, and  $L = 258$ ,  $K = 129$ ,  $s = 20$ ,  $\mathcal{C}_0 = \{-1, 0, +1\}$ . To ensure convergence, all algorithms perform 50 iterations. The measure of interest when dealing with discrete values is the symbol error rate  $\text{SER} = \frac{1}{L} \sum_{i=1}^L \Pr\{\hat{x}_i \neq x_i\}$ . The results are shown in Fig. 3. In the SVD-based case (upper figure), the new algorithm (yellow, dashed) performs as good as IMS/Q (red) [22]. TSR/Q

<sup>2</sup>For a detailed derivation, please see the Appendix.

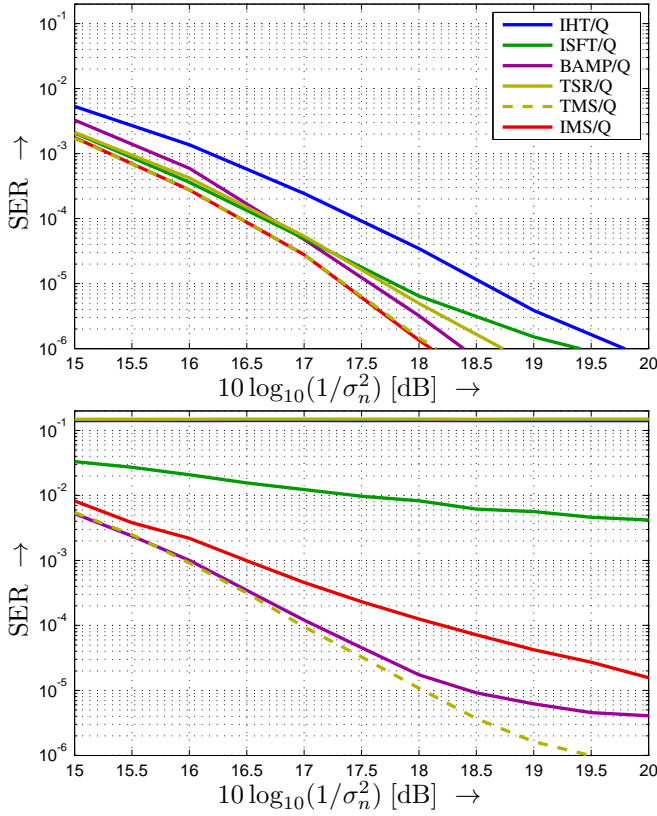


Fig. 3: SER over the noise level  $1/\sigma_n^2$  in dB.  $L = 258$ ,  $K = 129$ ,  $s = 20$ ,  $\mathcal{C} = \{-1, +1\}$ . SVD-based matrix (top), Gaussian matrix (bottom).  $\|\mathbf{A}_{(:,i)}\|^2 = 1 \forall i$ .

(yellow, solid) is outperformed by 0.5 dB, and the well-known BAMP/Q by 0.2 dB. For comparison, the results of the standard algorithms for discrete CS (IHT/Q, blue, and ISFT/Q, green) are also shown.

For Gaussian sensing matrices, the new algorithm TMS/Q shows the best performance. Note that TSR/Q fails since the assumption that  $\mathbf{A}$  is a part of an orthogonal matrix is not fulfilled. In general, the Gaussian matrices are less structured than the SVD-based, and only a smaller sparsity is tolerated to obtain the same performance. Note that the influence of the matrix is much larger for algorithms in which the matched filter is applied (IHT/Q, ISFT/Q) than for MMSE-estimation-based algorithms (TMS/Q, IMS/Q). However, also the aforementioned algorithms are able to find the correct result if the sparsity is small enough.

## V. CONCLUSION

In this paper, we have shown that the calculation of extrinsic information is equal to unbiasing, justifying the calculation performed in the second step of TSR and TMS. In addition, the TSR algorithm has been improved by the application of the LMMSE estimation, which leads to better numerical results than the established algorithms.

## REFERENCES

- [1] D.L. Donoho. Compressed Sensing. *IEEE Tr. Inf. Theory*, pp. 1289–1306, Apr. 2006.
- [2] H. Zhu, G.B. Giannakis. Exploiting Sparse User Activity in Multiuser Detection. *IEEE Tr. Comm.*, pp. 454–465, Feb. 2011.
- [3] R.F.H. Fischer, F. Wackerle. Peak-to-Average Power Ratio Reduction in OFDM via Sparse Signals: Transmitter-Side Tone Reservation vs. Receiver-Side Compressed Sensing. *Proc. Int. OFDM Workshop*, Essen, Germany, Aug. 2012.
- [4] A. Ens, A. Yousaf, T. Ostertag, L.M. Reindl. Optimized Sinus Wave Generation with Compressed Sensing for Radar Applications. *Proc. CoSeRa*, Bonn, Germany, Sept. 2013.
- [5] P. Dymarski, R. Romaniuk. Sparse Signal Modeling in a Scalable CELP Coder. *Proc. EUSIPCO*, Marrakech, Morocco, Sep. 2013.
- [6] R. Fay. Introducing the Counter Mode of Operation to Compressed Sensing Based Encryption. *Inf. Proc. Letters*, pp. 279–283, Apr. 2016.
- [7] Y.C. Pati, R. Rezaifar, P.S. Krishnaprasad. Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition. *Proc. Asilomar Conf.*, pp. 40–44, Nov. 1993.
- [8] T. Blumensath, M.E. Davis. Iterative Thresholding for Sparse Approximations. *Journal of Fourier Analysis and Appl.*, pp. 629–654, Dec. 2008.
- [9] I. Daubechies, M. Fornasier, I. Loris. Accelerated Projected Gradient Method for Linear Inverse Problems with Sparsity Constraints. *Journal of Fourier Analysis and Appl.*, pp. 764–792, Dec. 2008.
- [10] G.L. Nemhauser, L.A. Wolsey. *Integer and Combinatorial Optimization*, John Wiley & Sons, New York, 1988.
- [11] S. Sparrer, R.F.H. Fischer. Adapting Compressed Sensing Algorithms to Discrete Sparse Signals. *Proc. WSA*, Erlangen, Germany, Mar. 2014.
- [12] R.G. Baraniuk, V. Cevher, M.F. Duarte, C. Hedge. Model-Based Compressive Sensing. *IEEE Trans. Inf. Theory*, pp.1982–2001, Apr. 2010.
- [13] S. Sparrer, R.F.H. Fischer. Soft-Feedback OMP for the Recovery of Discrete-Valued Sparse Signals. *Proc. EUSIPCO*, Nice, France, Aug. 2015.
- [14] S. Sparrer, R.F.H. Fischer. An MMSE-Based Version of OMP for the Recovery of Discrete-Valued Sparse Signals. *Electronics Letters*, pp. 75–77, Jan. 2016.
- [15] D.L. Donoho, A. Maleki, A. Montanari. Message Passing Algorithms for Compressed Sensing: I. Motivation and Construction. *Proc. ITW*, Cairo, Egypt, Jan. 2010.
- [16] M. Bayati, A. Montanari. The Dynamics of Message Passing on Dense Graphs, with Application to Compressed Sensing. *IEEE Tr. Inf. Theory*, pp. 764–785, Feb. 2011.
- [17] F.R. Kschischang, B.J. Frey, H.-A. Loeliger. Factor Graphs and the Sum-Product Algorithm. *IEEE Tr. Inf. Theory*, pp. 498–519, Feb. 2001.
- [18] W. Dai, O. Milenkovic. Sparse Weighted Euclidean Superimposed Coding for Integer Compressed Sensing. *Proc. CISS*, pp. 470–475, Princeton, New Jersey, USA, March 2008.
- [19] W. Dai, O. Milenkovic. Weighted Euclidean Superimposed Codes for Integer Compressed Sensing. *Proc. ITW*, pp.124–128, Porto, Portugal, May 2008.
- [20] J. Ma, X. Yuan, L. Ping. Turbo Compressed Sensing with Partial DFT Sensing Matrix. *IEEE Signal Proc. Letters*, pp. 158–161, Feb. 2015.
- [21] J. Ma, X. Yuan, L. Ping. On the Performance of Turbo Signal Recovery with Partial DFT Sensing Matrices. *IEEE Signal Proc. Letters*, pp. 1580–1584, Oct. 2015.
- [22] S. Sparrer, R.F.H. Fischer. Algorithms for the Iterative Estimation of Discrete-Valued Sparse Vectors. *Proc. SCC*, Hamburg, Germany, Feb. 2017.
- [23] F. Tarköy. MMSE-Optimal Feedback and its Applications. *Proc. ISIT*, p. 334, Sep. 1995.
- [24] S. Sparrer, R.F.H. Fischer. Enhanced Iterative Hard Thresholding for the Estimation of Discrete-Valued Sparse Signals. *Proc. EUSIPCO*, Budapest, Hungary, Aug. 2016.
- [25] S.M. Kay. *Fundamentals of Statistical Signal Processing: I. Estimation Theory*, Prentice-Hall Inc., Upper Saddle River, NJ, USA, 1993.
- [26] C. Berrou, A. Glavieux, P. Thitimajshima. Near Shannon Limit Error-Correcting Coding and Decoding: Turbo-Codes (1). *Proc. ICC*, pp. 1064–1070, Geneva, Switzerland, May 1993.
- [27] R.F.H. Fischer. *Precoding and Signal Shaping for Digital Transmission*. John Wiley & Sons, Inc., New York, 2002.
- [28] Q. Guo, D.D. Huang. A Concise Representation for the Soft-In Soft-Out LMMSE Detector. *IEEE Comm. Letters*, pp. 566–568, May 2011.
- [29] P. Banelli. Non-Linear Transformations of Gaussians and Gaussian-Mixtures with Implications on Estimation and Information Theory. *arXiv*, 1111.5950v3, 10. May 2013.

## APPENDIX

In the following, we proof that  $E_{X,E}\{g(x+e) \cdot e\} = \sigma_{S,\text{post}}^2$ . Since we assume  $x$  and  $e$  in the channel model  $z = x + e$  to be independent, we have

$$\begin{aligned} E_{X,E}\{g(x+e) \cdot e\} &= \int \int g(x+e) e f_X(x) f_E(e) de dx \\ &= \int f_X(x) \left( \int (z-x) f_E(z-x) g(z) dz \right) dx \\ &\stackrel{\text{def}}{=} \int f_X(x) I_1(x) dx \end{aligned} \quad (23)$$

$$\begin{aligned} \sigma_{S,\text{post}}^2 &= E_Z\{\sigma_{S,\text{post},i}^2(z)\} = E_{X,E}\{\sigma_{S,\text{post},i}^2(x+e)\} \\ &= \int \int \sigma_{S,\text{post},i}^2(x+e) f_X(x) f_E(e) de dx \\ &= \int f_X(x) \left( \int f_E(z-x) \sigma_{S,\text{post},i}^2(z) dz \right) dx \\ &\stackrel{\text{def}}{=} \int f_X(x) I_2(x) dx . \end{aligned} \quad (24)$$

In order to proof that  $E\{e \cdot g(x+e)\} = \sigma_{S,\text{post}}^2$  holds, we show that  $I_1(x) = I_2(x)$ , which is a stricter condition than to proof “only” the equality of the entire integrals.

In the following, we replace the variable  $x$  of the integral by  $q$  to avoid confusion. With  $g(z) = E\{x|z\} = \int x f_X(x|z) dx$ ,  $\sigma_{S,\text{post},i}^2 = E\{x^2|z\} - (E\{x|z\})^2 = \int x^2 f_X(x|z) dx - (\int x f_X(x|z) dx)^2$ ,  $f_Z(z|x) = f_E(z-x)$ , and Bayes’ theorem follows<sup>3</sup>

$$\begin{aligned} I_1(q) &= \int (z-q) f_E(z-q) \frac{\int x f_E(z-x) f_X(x) dx}{f_Z(z)} dz \\ I_2(q) &= \int f_E(z-q) \frac{\int x^2 f_E(z-x) f_X(x) dx}{f_Z(z)} dz \\ &\quad - \int f_E(z-q) \left( \frac{\int x f_E(z-x) f_X(x) dx}{f_Z(z)} \right)^2 dz . \end{aligned}$$

In the following, we assume that the error is Gaussian distributed with variance  $\sigma_e^2$ , i.e.,  $f_E(e) = c e^{-e^2/(2\sigma_e^2)}$ ,  $c = \frac{1}{\sqrt{2\pi\sigma_e^2}}$ . We reformulate  $I_1(q)$  via integration in parts. It holds

$$\begin{aligned} \int f'(z)g(z) dz &= [f(z)g(z)]_{-\infty}^{\infty} - \int g'(z)f(z) dz \\ &= - \int g'(z)f(z) dz , \end{aligned} \quad (25)$$

with

$$f'(z) = (z-q) f_E(z-q) = (z-q) c e^{-(z-q)^2/(2\sigma_e^2)} \quad (26)$$

$$f(z) = -\sigma_e^2 c e^{-(z-q)^2/(2\sigma_e^2)} \quad (27)$$

$$g(z) = \frac{\int x f_E(z-x) f_X(x) dx}{f_Z(z)} \stackrel{\text{def}}{=} \frac{u(z)}{v(z)} \quad (28)$$

$$g'(z) = \frac{u'(z)v(z) - u(z)v'(z)}{(v(z))^2} . \quad (29)$$

<sup>3</sup>If no limits are given for integrals, the lower and upper limits are  $-\infty$  and  $\infty$ , respectively.

Note that  $[f(z)g(z)]_{-\infty}^{\infty} = 0$  since  $\lim_{|z| \rightarrow \infty} f(z) = 0$ . For Gaussian error, the factors in (29) are given by

$$\begin{aligned} u(z) &= \int x f_E(z-x) f_X(x) dx \\ &= \int x c e^{-(z-x)^2/(2\sigma_e^2)} f_X(x) dx \\ u'(z) &= \frac{d}{dz} \int x c e^{-(z-x)^2/(2\sigma_e^2)} f_X(x) dx \\ &= - \int x c \frac{1}{\sigma_e^2} (z-x) e^{-(z-x)^2/(2\sigma_e^2)} f_X(x) dx \\ &= - \frac{1}{\sigma_e^2} \int x (z-x) f_E(z-x) f_X(x) dx \\ v(z) &= f_Z(z) \\ &= \int f_E(z-x) f_X(x) dx \\ &= \int c e^{-(z-x)^2/(2\sigma_e^2)} f_X(x) dx \\ v'(z) &= \frac{d}{dz} \int c e^{-(z-x)^2/(2\sigma_e^2)} f_X(x) dx \\ &= - \int c \frac{1}{\sigma_e^2} (z-x) e^{-(z-x)^2/(2\sigma_e^2)} f_X(x) dx \\ &= - \frac{1}{\sigma_e^2} \int (z-x) f_E(z-x) f_X(x) dx \\ &= - \frac{1}{\sigma_e^2} z \int f_E(z-x) f_X(x) dx \\ &\quad + \frac{1}{\sigma_e^2} \int x f_E(z-x) f_X(x) dx . \end{aligned}$$

Hence  $g'(z)$  calculates to

$$\begin{aligned} g'(z) &= \frac{u'(z)v(z) - u(z)v'(z)}{(v(z))^2} \\ &= - \frac{1}{(v(z))^2} \cdot \frac{1}{\sigma_e^2} \cdot \\ &\quad \left( z \cdot \int x f_E(z-x) f_X(x) dx \cdot \int f_E(z-x) f_X(x) dx \right. \\ &\quad - \int x^2 f_E(z-x) f_X(x) dx \cdot \int f_E(z-x) f_X(x) dx \\ &\quad - z \cdot \int f_E(z-x) f_X(x) dx \cdot \int x f_E(z-x) f_X(x) dx \\ &\quad \left. + \int x f_E(z-x) f_X(x) dx \cdot \int x f_E(z-x) f_X(x) dx \right) \\ &= \frac{1}{(v(z))^2} \cdot \frac{1}{\sigma_e^2} \cdot \\ &\quad \left( \int x^2 f_E(z-x) f_X(x) dx \cdot \int f_E(z-x) f_X(x) dx \right. \\ &\quad \left. - \int x f_E(z-x) f_X(x) dx \cdot \int x f_E(z-x) f_X(x) dx \right) \end{aligned} \quad (30)$$

Plugging (26)–(28) and (30) into (25) yields

$$I_1(q) = - \int f(z) g'(z) dz$$

$$\begin{aligned}
&= \int \sigma_e^2 c e^{-(z-q)^2/(2\sigma_e^2)} \cdot \frac{1}{(\mathbf{f}_Z(z))^2 \cdot \sigma_e^2} \cdot \\
&\quad \left( \int x^2 \mathbf{f}_E(z-x) \mathbf{f}_X(x) dx \cdot \int \mathbf{f}_E(z-x) \mathbf{f}_X(x) dx \right. \\
&\quad \left. - \left( \int x \mathbf{f}_E(z-x) \mathbf{f}_X(x) dx \right)^2 \right) dz \\
&= \int \mathbf{f}_E(z-q) \cdot \left( \frac{\int x^2 \mathbf{f}_E(z-x) \mathbf{f}_X(x) dx}{\mathbf{f}_Z(z)} \right) dz \\
&\quad - \int \mathbf{f}_E(z-q) \cdot \left( \frac{\int x \mathbf{f}_E(z-x) \mathbf{f}_X(x) dx}{\mathbf{f}_Z(z)} \right)^2 dz \\
&= I_2(q) .
\end{aligned}$$

Thus, both integrals are equal, and hence  $\mathbb{E}_{X,E}\{g(x+e) \cdot e\} = \sigma_{S,\text{post}}^2$ .

Note that the unbiasing is independent of the distribution  $\mathbf{f}_X(x)$ , i.e., it holds for all possible alphabets, hence for discrete, and also for continuous (real-valued) distributions of the elements of the sparse vector. The only condition is that the *error* is Gaussian distributed.